DOCUMENT RESUME

ED 164 612 TM 008 181

AUTHOR Dawson, Beth K.
TITLE Bias of the Canonical Redundancy Statistic: A Monte Carlo Study.
PUB DATE Mar 78
NOTE 29p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978) ; Best copy available

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS *Correlation; Evaluation Criteria; *Multiple Regression Analysis; Predictor Variables; *Statistical Bias
IDENTIFIERS *Canonical Correlation

ABSTRACT

A Monte Carlo study was conducted to discover the amount of bias to be expected in canonical correlations and in the canonical redundancy statistic. Six hundred random sample replications of 72 conditions were performed and the mean sample values compared with population values. Conditions were varied with respect to the number of predictor variables, the number of criterion variables, the size of the population correlations, and the size of the samples. The least biased estimates are those that use large sample sizes with moderate to high interbattery correlations. Both the Wherry and the Olkin-Pratt correction formulas were found to be satisfactory for correcting for bias in the total redundancy statistic. (Author/CTM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED164612

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BIAS OF THE CANONICAL REDUNDANCY STATISTIC

A Monte Carlo Study

Beth K. Dawson
Southern Illinois University School of Medicine

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Beth Dawson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

BEST COPY AVAILABLE

Presented at
The Annual Meeting of the American Educational Research Association

March, 1978
Toronto, Canada

TM008 181

BIAS OF THE CANONICAL REDUNDANCY STATISTIC:

A Monte Carlo Study

Canonical correlation analysis is a multivariate statistical procedure used to study the relationship between two sets of variables. The squared canonical correlation coefficients describe the amount of variance accounted for by successive pairs of linear combinations of predictor variables and linear combinations of criterion variables. A new measure, the canonical redundancy statistic, has been proposed (Stewart and Love, 1968, and Miller, 1969) as an estimate of the amount of shared variance of the variables themselves and, as such, is an appealing alternative to the squared canonical correlation coefficient. Instead of one measure of association, the redundancy statistics provide a separate estimate for each set of variables, given knowledge of the other set.

The sampling distributions of various correlational statistics have been amply investigated and it is well known that these measures are biased. Of particular interest in multivariate analysis are the multiple correlation coefficient and the canonical correlation coefficients. No investigation has been reported on the sampling distribution and bias of the canonical redundancy statistic. Yet it is almost a certainty that this statistic is biased: the mathematical definition illustrates that it may be described as a linear combination of the canonical correlation coefficients. Thus the purpose of the present study is to investigate the bias of the canonical redundancy statistic, especially as compared to that of the canonical correlation coefficient, and to develop a formula to estimate the true population value on the basis of sample information.

The basis for canonical correlation analysis was developed by Hotelling (1935, 1936). He defined "the most predictable criterion" as the linear combination of criterion variables that is predicted by a linear combination of predictor variables so that the two linear combinations have the highest possible correlation.

When the influence of the first two linear combinations is partialled out, the process is repeated on the residuals, thus obtaining a sequence of pairs of variates with maximum correlations between them. These were denoted by Hotelling as canonical variates and canonical correlations, respectively.

Many authors subsequently noted that determination of the "most predictable criterion" is not always the appropriate goal of educational or psychological research. However, it was immediately appreciated that the technique developed by Hotelling provides a mechanism to study the number and nature of mutually independent relations between two sets of variables.

Darlington, Weinberg and Walberg (1973) described the manner in which canonical correlation analysis assists in researching such relationships. First, it determines the minimum number of traits needed to account for the important linear relationships between two batteries. For example, a researcher might hypothesize that there are r traits that describe the important relationships between a set of attitude variables and a set of performance variables. After performing a canonical correlation analysis, the number of significant relationships may be determined by testing whether the canonical correlation coefficient is greater than zero. Second, the standardized weights and factor structures assist in describing the nature of these traits.

Theoretical research in canonical correlation analysis has included the generalization to the situation in which more than two sets of variables may be analyzed (Horst, 1961) and an assessment of the effects of unreliability of measures on a canonical correlation solution (Meredith, 1964). In the area of applications, many investigators have discussed the use of standardized coefficients versus factor structure correlations to facilitate interpretation of the canonical variates (Tatsuoka, 1973b; Cooley and Lohnes, 1971; Barcikowski and Stevens, 1975). The stability of canonical weights and correlations under various conditions were studied by Kau

(1976), and Rim (1972) investigated the shrinkage of the canonical correlation on cross-validation.

Practical problems of interpretation of canonical correlation results led to the simultaneous development of the canonical redundancy statistic by Stewart and Love (1968) and Miller (1969). Both demonstrated the equivalence between this statistic and the average squared multiple correlation obtained from the regression of each variable in one set on all variables in the other set. Although the basic definition of the redundancy statistic was questioned by Nicewander and Wood (1974, 1975), the recent work by Gleason (1976) has established the mathematical basis for the statistic.

An outline of the eigenanalysis procedure used in canonical correlation and the definition of the redundancy statistic follows:

Let $\underline{R}_{xx}$ be the matrix of correlations between p predictor variables;

$\underline{R}_{yy}$ be the matrix of correlations between q criterion variables;

$\underline{R}_{xy}$ be the matrix of correlations between the predictor and criterion variables.

The canonical correlation solution is obtained from eigenanalysis of the nonsymmetric matrix formed by the product $\underline{R}_{xx}^{-1}\underline{R}_{xy}\underline{R}_{yy}^{-1}\underline{R}_{yx}$.

Let Λ be the resulting diagonal matrix of eigenvalues with the ith diagonal element denoted λ_i^2;

$\underline{A}$ be the resulting matrix of eigenvectors for the predictor variables with the ith column denoted $\underline{a}_i$;

$\underline{B}$ be the resulting matrix of eigenvectors for the criterion variables resulting from eigenanalysis of $\underline{R}_{yy}^{-1}\underline{R}_{yx}\underline{R}_{xx}^{-1}\underline{R}_{xy}$ with the ith column denoted $\underline{b}_i$.

Then $v_{xi} = \underline{a}_i'\underline{x} = a_{i_1}x_1 + a_{i_2}x_2 + \cdots + a_{ip}x_p$ is the ith canonical variate for the predictor variables, i=1, ..., s, where s=min(p,q);

$v_{yj} = \underline{b}_j'\underline{y} = b_{j_1}y_1 + b_{j_2}y_2 + \cdots + b_{jq}y_q$ is the jth canonical variate for the criterion variables, j=1, ..., s.

The eigenvalues are the squared canonical correlation coefficients between each successive pair of canonical variates. The eigenvector matrices, $\underline{A}$ and $\underline{B}$, contain the standardized weights for the predictor and criterion variables, respectively, and are used to form the canonical variates. If eigenanalysis is performed using variance-covariance matrices in place of correlation matrices the eigenvalue matrix must be postmultiplied by the diagonal matrix of standard deviations of the appropriate set of variables, predictor or criterion, to obtain the standardized weights.

The canonical redundancy statistic is defined as the sum of successive products between the proportion of variance that the canonical variates of either battery explain in the canonical variates of the other (accounted for by the squared canonical correlations), and the proportion of variance absorbed by the canonical variates from their respective batteries (accounted for by the squared correlations between the variates and original variables). Using the above notation, the total canonical redundancy statistic for the predictor variables, given the criterion variables, is

$$Rd_x = \frac{1}{p} \Sigma_j^s (\underline{R}_{xx}\underline{a}_j)' (\underline{R}_{xx}\underline{a}_j) \lambda_j^2;$$

the total canonical redundancy statistic for the criterion variables, given the predictor variables, is similarly

$$Rd_y = \frac{1}{q} \Sigma_j^s (\underline{R}_{yy}\underline{b}_j)' (\underline{R}_{yy}\underline{b}_j) \lambda_j^2;$$

where s = min(p,q).

This sum determines the amount of redundancy in one battery of variables given the other. As such, it is directional and nonsymmetric and has a desirable range of zero to one.

The development of an index that describes the overlap or amount of redundancy between two batteries in canonical correlation analysis was extremely welcomed by researchers in education and psychology. Since the papers by Stewart and Love (1968) and Miller (1969) were published, the canonical redundancy statistic has been described and recommended in texts and articles by some of social sciences' leading authors. For examples, see Cooley and Lohnes (1971, p. 170-172); Tatsuoka (1973b, p. 280-282); Cohen and Cohen (1975, p. 429-432); Timm (1975, p. 355-358) and Cooley and Lohnes (1976, p. 211-212). In addition, an entire session at the 1976 Annual Meeting of the American Education Research Association was devoted to applied research on the redundancy statistic.

METHOD

It is almost a certainty that the redundancy statistic is positively biased. First, consider the dependence of this statistic on the squared canonical correlation coefficient. Second, the results by Miller (reported in Cooley and Lohnes, 1976) of a Monte Carlo analysis investigating the sampling distribution in the null case indicate bias of the median total redundancy statistic ranging from .06 to .09 for various combinations of numbers of predictor and criterion variables and sample sizes. No other data are available on the bias of this statistic. Thus the purpose of the present study was to investigate the empirical sampling distribution of the first squared canonical correlation coefficient and of the redundancy statistic using Monte Carlo methods; subsequently, an attempt was made to derive a formula to correct for bias in the total redundancy statistic.

The investigation entailed the systematic variation of the number of predictor and criterion variables, the sample size, the size of the intrabattery correlations

and the size of the interbattery correlations since these are the six factors that affect the magnitude of the canonical redundancy statistic.

Two levels of the number of variables were designated for each of the left and right sets: cases with five and nine variables for each set. Due to the summetry of canonical correlation analysis, it was necessary to consider only the following three combinations:

$$\{(p,q)\} = \{(5,5), (9,5),(9,9)\}.$$

Two sample sizes were selected for study, the case of a small sample and the case of a large sample. For small n, the value used was 5(p+q), a sample size frequently encountered in applied research. For the large sample, n equal to 20(p+q) was used to examine effects of a sample size frequently recommended.

Two conditions were chosen for the average intracorrelations of each of the matrices $\underset{\sim}{P}_{xx}$ and $\underset{\sim}{P}_{yy}$, and three conditions were selected for $\underset{\sim}{P}_{xy}$. In applied research, variables in the predictor and criterion sets often have medium to high correlations. However, the correlations between predictor and criterion variables are quite often lower. In an attempt to reflect conditions often encountered in actual research, the off-diagonal elements of $\underset{\sim}{P}_{xx}$ and $\underset{\sim}{P}_{yy}$ were set to .30 and .60 to reflect medium and high correlations respectively. All of the elements of $\underset{\sim}{P}_{xy}$ were set to .00, .20 and .40 to reflect the null case, low correlations and medium correlations, respectively. The inclusion of the null case for no relationship between the two sets of variables was important in this study to provide baseline information against which to compare bias in the non-null case.

The above conditions lead to the definition of 36 population matrices. Calculation of parameters and statistics based upon 2 sample sizes increases the number of specific situations under investigation to 72.

A computer algorithm developed by Montanelli (1971) using the Odell-Fieveson method of generating sample correlation matrices was utilized to generate a supermatrix representing each of the 72 different population situations. These generated supermatrices were then defined as population matrices and were subjected to canonical and redundancy analysis. Subsequently, each of the 72 newly defined population matrices were used to generate 600 sample matrices that were, in turn, eigenanalyzed to calculate sample estimates of the statistics under investigation. In each case, the 600 sample estimates were aggregated in order to compare their values with the population values calculated on the corresponding population matrix.

The following five statistics were calculated.

R_1^2: the squared canonical correlation coefficient between the first pair of canonical variates;

Rd_{x_1}, Rd_{y_1}: the canonical redundancy statistics for each set of variables, predictor and criterion, corresponding to the first pair of canonical variates;

Rd_x, Rd_y: the total canonical redundancy statistics for each set of variables corresponding to all pairs of canonical variates.

The mean, standard deviation and the average bias of the five statistics were determined for each of the 72 population situations. See Figure 1 for a schematic design of the Monte Carlo procedure.

RESULTS

Results of the Monte Carlo analysis indicate that the redundancy statistics are less biased than the first squared canonical correlation coefficient, both for large and small samples (Table 1). The most important parameter affecting the bias of all statistics is the sample size: bias increases fourfold as n increased from 5(p+q) to 20(p+q). The sampling standard deviations of the statistics double in the same situation. It is not known whether this pattern is consistent across sam-

ple sizes not considered in this study. The second most critical parameter affecting the amount of bias is the level of interbattery correlations: in all cases bias decreases as interbattery correlations increase, indicating less bias in the non-null case. The decrease is less marked for the total redundancy statistic.

Next in importance is the number of predictor and criterion variables and here the statistics demonstrate an interesting behavior. While bias of R_1^2 increases as both p and q increase, bias of Rd_x decreases and that of Rd_y increases as p increases from 5 to 9 and q remains at 5. The limited number of levels of p and q examined in the present study does not permit the generalization that bias of redundancies defined on the battery with a large number of variables is smaller, but this is an interesting pattern deserving further investigation. Bias of all the statistics is least affected by the level of intrabattery correlations.

Table 2a examines the mean bias of these statistics for varying numbers of prediction and criterion variables and intrabattery correlations when sample sizes are fixed and interbattery correlations are zero, the null case. Table 2b presents identical data for the case when $\underline{R}_{xy}$=.40. In the null case, it can be seen that the bias of R_1^2 increases as p and q increase for both small and large sample sizes. This pattern is absent for the redundancy statistics. All of the statistics demonstrate minimal variation in bias within given levels of p and q, regardless of the level of intrabattery correlations, a further indication of the minimal effect of this factor.

Table 2b illustrates the dramatic decrease in bias as interbattery correlation increases. Again, bias of R_1^2 is more affected by levels of p and q than is that of the redundancy statistics. The bias of the redundancies associated with the first pair of canonical variates and that of the total redundancies for large samples is small enough to be little or no concern. Within levels of p and q, bias tends to increase as intrabattery correlations increase, especially in the small sample case.

The increase is significant for R_1^2 but only about .02 to .03 for the total redundancy statistics.

The degree of bias is related to the size of the population parameter due to the restricted range [0, 1] of these statistics. Furthermore, a bias of .05 is of less concern in a situation in which the population value is .50 than that in which the population value is .10. Consequently, a coefficient of bias was defined as the ratio of the observed mean of the estimate to the estimate of the population value. An unbiased estimate would thus have a coefficient of bias equal to unity.

Table 3 presents values of the coefficient of bias for each of the statistics collapsed across various factors of p, q, etc. It can be seen that the coefficient of bias is, in general, fairly similar for the first squared canonical correlation coefficient and the total redundancy statistics. The coefficient of bias is least effected by levels of intrabattery correlation but is seen to decrease in the non-null case and for large samples. The behavior of the coefficient of bias when sample size and intrabattery correlations are held constant is similar to that of mean bias.

The results of the Monte Carlo Analyses indicate that the behavior of the bias of total redundancy statistic is quite similar to that of the squared canonical correlation coefficient. In addition, Miller (1975b, 1976) demonstrated that the redundancy statistics are approximated by an F distribution with modified degrees of freedom in the null case. It was therefore decided to apply the Wherry and the Olkin-Pratt (Kendall and Stuart, 1967) shrinkage formulae for the squared multiple correlation coefficient in the hope that the population values of the redundancy statistics might be similarly estimated.*

*The author would like to express her appreciation to John Pohlmann, Southern Illinois University, for his suggestion to examine the efficacy of these formulae.

The formula used were the following:

Let N be the sample size,

q be the number of variables in the criterion set,

Rd be the total redundancy statistic for the predictor set given the criterion set,

Then

$$\text{Wherry correction} = 1 - \frac{N-1}{N-q-1}\ (1-Rd);$$

$$\text{Olkin-Pratt correction} = 1 - \frac{N-3}{N-q-1}\ (1-Rd) - \left(\frac{N-3}{N-q-1}\right)\left(\frac{2}{N-q+1}\right)(1-Rd)^2.$$

In the case of the total redundancy for the criterion set given the predictor set, the roles of p and q were exchanged to be consistent with the above.

Each formula was applied to the mean sample values of the total redundancy statistic and the difference between the population value and the shrunken estimate was calculated. In addition, the mean absolute value of the difference was computed over all 144 values of Rd. Tables 4a and 4b contain the results of this analysis for the cases in which p=9 and q=5.

Both formulae provide excellent approximations to the population value. The average absolute values of the difference between the population and the corrected value were 0.003347 for the Wherry formula and 0.001955 for the Olkin-Pratt formula. Although the Olkin-Pratt formula is better when considering only the residuals, Tables 4a and 4b illustrate that this formula results in more overestimates of the population values. In the complete analysis, this is evidenced by the larger number (82) of negative differences as compared to only 12 overestimates using the Wherry formula. With the Wherry formula, only 3 of the 144 estimates vary from the population parameter by a value greater than 0.01, and these are all underestimating the parameter.

SUMMARY

Comparison of Tables 1 and 3 indicates that, although the amount of bias is

larger for the total redundancy statistics than for the redundancies corresponding to the first canonical variate, the coefficients of bias are similar in magnitude for these two measures. The consequence for applied researchers is that they may examine the redundancies based upon a larger number of canonical variates without increasing the proportion of bias that would be present when interpreting redundancy based upon only the first two or three variates.

The least biased estimates, for both R_1^2 and Rd_x and Rd_y, are those that use large sample sizes with moderate to high levels of interbattery correlations. The researcher may be less concerned with the number of variables in a study, except as this number relates to sample size, and with the degree of relationship among the variables in each individual battery.

Two commonly used formulae developed to predict the true population value of the squared multiple correlation coefficient are found to be satisfactory for the total redundancy statistic. These are the Wherry and Olkin-Pratt correction formula (Kendall and Stuart, 1967). The Wherry formula, while slightly less accurate than Olkin-Pratt, overestimates the true population value in only about eight percent of the cases. Applying this formula to the mean sample estimates, the corrected values differ from their true counterparts by no more than 0.01, and again, this is an underestimate. Thus, the Wherry formula is recommended for its tendency to provide an accurate yet slightly conservative estimate of the true population parameter.

It would seem that the results of the Monte Carlo analysis do not justify the recommendation of the canonical redundancy statistic as an alternative to the squared canonical correlations on the basis of less bias. The redundancy statistic, in general, appears to exhibit a degree of bias quite similar to that of the squared canonical correlation. However, the bias is easily corrected by the Wherry or Olkin-Pratt formulae to estimate the true population value. What is perhaps

more important is the interpretive characteristics of the canonical redundancy statistic as compared to those of the canonical correlation coefficient. If interest truly lies in the relationship between groups of variables as opposed to the relationship between linear combinations of variables, then the redundancy statistic provides a more realistic and meaningful measure for the conscientious educational researcher.

References

Anderson, T.W. An Introduction of Multivariate Statistical Analysis. New York: Wiley, 1958.

Barcikowski, R.S. and Stevens, J.P. A Monte Carlo study of the stability of canonical correlations, canonical weights and canonical variate-variable correlations. Multivariate Behavioral Research, 1975, 10: 353-364.

Cohen, J. and Cohen, P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. New York: Wiley, 1975.

Cooley, W.W. and Lohnes, P.R. Multivariate Data Analysis. New York: Wiley, 1971.

Cooley, W.W. and Lohnes, P.R. Evaluation Research in Education. New York: Wiley, 1976.

Darlington, R.B., Weinberg, S.L. and Walberg, H.J. Canonical variate analysis and related techniques. Review of Educational Research, 1973, 43: 433-454.

Gleason, T.C. On redundancy in canonical analysis. Psychological Bulletin, 1976, 83: 1004-1006.

Horst, P. Relations among m sets of measures. Psychometrika, 1961, 26: 129-149.

Hotelling, H. The most predictable criterion. Journal of Educational Psychology, 1935, 26: 130-142.

Hotelling, H. Relations between two sets of variates. Biometrika, 1936, 28: 321-377.

Kau, D.C. A Monte Carlo study of the estimation of canonical regression coefficients and correlations. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1976.

Kendall, M.G. and Stuart, A. The Advanced Theory of Statistics (2nd Ed.), Volume 2. New York: Harper, 1967.

Lohnes, P.R. and Cooley, W.W. Partitioning redundancy among predictor domains in reduced-rank models for school effects. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1976.

Meredith, W. Canonical correlations with fallible data. Psychometrika, 1964, 29: 55-65.

Miller, J.K. The development and application of bi-multivariate correlation. Unpublished doctoral dissertation, State University of New York at Buffalo, 1969.

Miller, J.K. In defense of the general canonical correlation index: Reply to Nicewander and Wood. Psychological Bulletin, 1975a, 82: 207-209.

Miller, J.K. The sampling distribution and a test for the significance of the bi-multivariate redundancy statistic: A Monte Carlo study. Multivariate Behavioral Research, 1975b, 10: 233-244.

Miller, J.K. Testing hypotheses about the bimultivariate redundancy statistic. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1976.

Montanelli, R.G., Jr. An investigation of the goodness of fit of the maximum likelihood estimation procedure in factor analysis. Unpublished doctoral dissertation, University of Illinois, Urbana, 1971.

Nicewander, W.A. and Wood, D.A. Comments on "A general canonical correlation index." Psychological Bulletin, 1974, 81: 92-94.

Nicewander, W.A. and Wood, D.A. On the mathematical bases of the general canonical correlation index: Rejoinder to Miller. Psychological Bulletin, 1975, 82: 210-212.

Odell, P.L. and Fieveson, A.H. A numerical procedure to generate a sample covariance matrix. Journal of the American Statistical Association, 1966, 61: 199-203.

Rim, Eui-do. A stepwise canonical approach to the selection of 'kernal' variables from two sets of variables. Unpublished doctoral dissertation, University of Illinois, Urbana, 1972.

Stewart, D. and Love, W. A general canonical correlation index. Psychological Bulletin, 1968, 70: 160-163.

Tatsuoka, M.M. Multivariate Analysis. New York: Wiley, 1971.

Tatsuoka, M.M. Multivariate analysis in educational research. In F.N. Kerlinger (Ed.), Review of Research in Education. Itaska, IL: Peacock, 1973b.

Timm, N.H. Multivariate Analysis with Applications in Education and Psychology. Monterey, CA: Brooks/Cole, 1975.

FIGURE 1

SCHEMATIC REPRESENTATION OF DESIGN OF MONTE CARLO ANALYSIS

Step 1: Define population conditions by specifying levels for p,q, etc.

Step 2: Form the required supermatrix for each of the 72 population conditions specified in Step 1.

Step 3: Use Montanelli procedure to generate a population supermatrix reflecting the specified population conditions.

Step 4: Perform eigenanalysis and calculate population values of each of the parameters under investigation.

Step 5: Use Montanelli procedure to generate 600 sample supermatrices from the population matrix obtained in Step 3.

Step 6: Perform eigenanalysis on each sample and calculate sample estimates of the parameters.

Step 7: Use the values from the samples to calculate the mean, standard deviation, average bias and coefficient of bias for each of the estimates.

Step 8: Compare the aggregated sample results with the original population parameters.

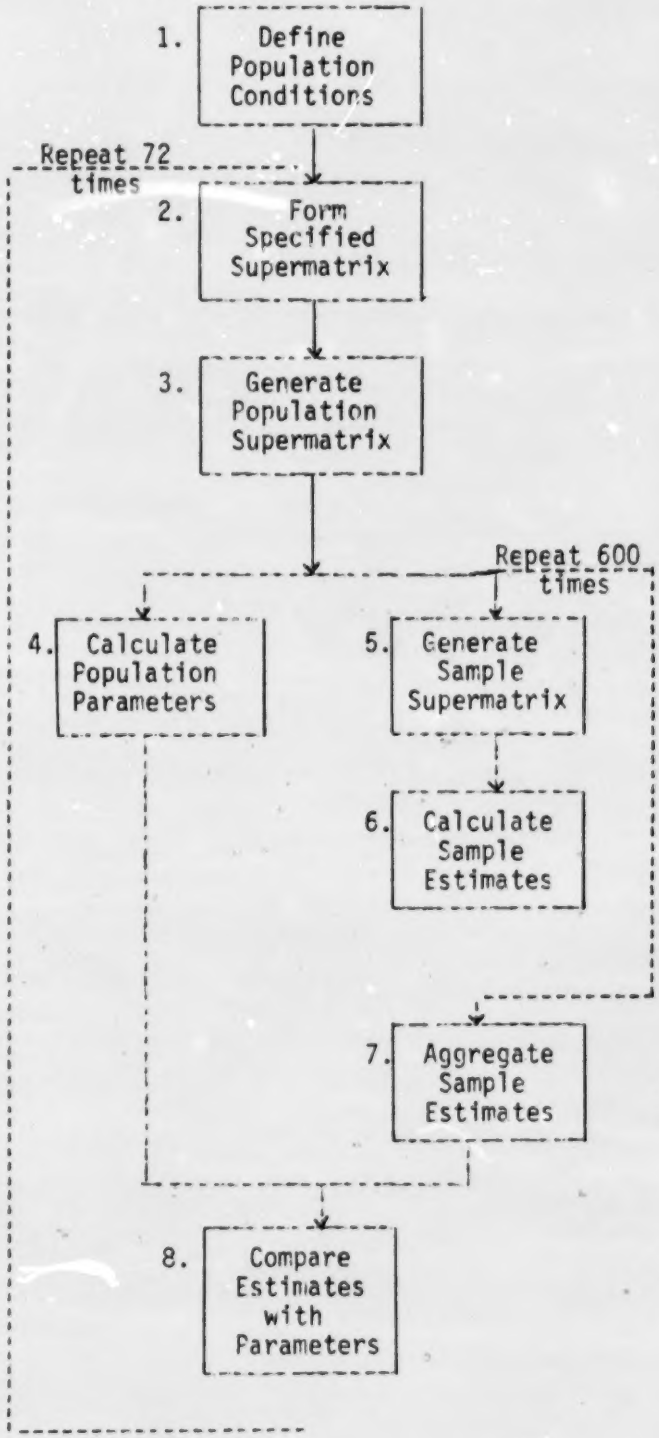

TABLE 1

COLLAPSED VALUES OF MEAN BIAS

COLLAPSING ACROSS ALL CONDITIONS EXCEPT:						
NUMBER OF VARIABLES		R_1^2	Rd_{x_1}	Rd_{y_1}	Rd_x	Rd_y
Predictor	Criterion					
5	5	.088	.022	.017	.049	.048
9	5	.095	.016	.018	.034	.058
9	9	.098	.013	.009	.045	.047
CORRELATION* BETWEEN PREDICTOR VARIABLES						
.30		.084	.017	.014	.042	.049
.60		.104	.017	.016	.044	.053
CORRELATION* BETWEEN CRITERION VARIABLES						
.30		.086	.015	.015	.041	.050
.60		.102	.019	.014	.044	.052
CORRELATION* BETWEEN PREDICTOR AND CRITERION VARIABLES						
.00		.121	.031	.023	.045	.055
.20		.111	.014	.015	.047	.056
.40		.049	.006	.006	.036	.042
SAMPLE SIZE**						
Small		.147	.025	.024	.069	.082
Large		.041	.009	.005	.017	.020

* Correlation in the matrix used to generate the population matrix analyzed in the study

** Small =5(p+q); Large = 20(p+q) where p is number of predictor variables, q is the number of criterion

TABLE 2a

VALUES OF MEAN BIAS FOR INTERBATTERY CORRELATIONS APPROXIMATING .00

				R_1^2		Rd_{X_1}		Rd_{Y_1}		Rd_X		Rd_Y	
p	q	Rxx*	Ryy*	ns**	nℓ**	ns**	nℓ**	ns**	nℓ**	ns**	nℓ**	ns**	nℓ**
5	5	.30	.30	.147	.033	.034	.007	.034	.008	.084	.020	.083	.020
5	5	.30	.60	.147	.032	.039	.007	.035	.008	.081	.019	.080	.019
5	5	.60	.30	.149	.033	.028	.005	.032	.008	.082	.018	.081	.021
5	5	.60	.60	.156	.033	.031	.005	.035	.007	.086	.020	.080	.020
9	5	.30	.30	.192	.068	.047	.029	.046	.015	.058	.015	.102	.025
9	5	.30	.60	.197	.068	.047	.028	.053	.019	.058	.015	.105	.026
9	5	.60	.30	.195	.071	.057	.040	.045	.015	.060	.015	.100	.026
9	5	.60	.60	.194	.066	.056	.037	.051	.019	.057	.013	.101	.024
9	9	.30	.30	.213	.066	.039	.017	.027	.007	.078	.018	.081	.020
9	9	.30	.60	.214	.066	.040	.018	.024	.005	.079	.018	.085	.021
9	9	.60	.30	.215	.067	.047	.021	.027	.007	.074	.016	.080	.018
9	9	.60	.60	.213	.064	.049	.020	.027	.005	.076	.016	.083	.018

* Correlation in the matrix used to generate the population matrix analyzed in the study
** Small = 5(p+q); Large = 20(p+q) where p is number of predictor variables, q is the number of criterion

TABLE 2b

VALUES OF MEANS BIAS FOR INTERBATTERY CORRELATIONS APPROXIMATING .40

				R_1^2		Rd_{x_1}		Rd_{y_1}		Rd_x		Rd_y	
p	q	Rxx*	Ryy*	ns**	nℓ**	ns**	nℓ**	ns**	nℓ**	ns**	nℓ**	ns**	nℓ**
5	5	.30	.30	.028	.006	.009	.003	.009	.002	.053	.013	.057	.012
5	5	.30	.60	.097	.023	.022	.006	.010	.005	.071	.017	.062	.015
5	5	.60	.30	.117	.022	.016	.005	.021	.004	.072	.017	.076	.016
5	5	.60	.60	.177	.047	.045	.016	.014	-.002	.081	.020	.076	.019
9	5	.30	.30	.005	.001	.006	.002	.006	.002	.039	.010	.061	.014
9	5	.30	.60	.058	.014	.008	.002	.013	.004	.046	.012	.063	.016
9	5	.60	.30	.073	.019	-.005	.003	.019	.006	.036	.012	.082	.022
9	5	.60	.60	.137	.035	-.013	-.005	.005	-.000	.050	.013	.090	.023
9	9	.30	.30	.000	.000	.002	.002	.005	.002	.047	.014	.057	.015
9	9	.30	.60	.059	.013	.009	.002	.008	.001	.064	.015	.054	.012
9	9	.60	.30	.059	.014	.006	.003	.010	.003	.051	.013	.067	.017
9	9	.60	.60	.146	.034	-.009	-.006	-.003	-.000	.074	.019	.074	.018

* Correlation in the matrix used to generate the population matrix analyzed in the study
** ns = 5(p+q); nℓ = 20(p+q) where p is number of predictor variables, q is the number of criterion

TABLE 3

COLLAPSED VALUES OF COEFFICIENT OF BIAS

COLLAPSING ACROSS ALL CONDITIONS EXCEPT:						
NUMBER OF VARIABLES		R_1^2	Rd_{x_1}	Rd_{y_1}	Rd_x	Rd_y
Predictor	Criterion					
5	5	1.246	1.371	1.198	1.349	1.512
9	5	1.244	1.576	1.291	1.212	1.686
9	9	1.266	1.551	1.238	1.238	1.138
CORRELATION* BETWEEN PREDICTOR VARIABLES						
.30		1.222	1.398	1.212	1.243	1.461
.60		1.282	1.600	1.273	1.289	1.430
CORRELATION BETWEEN CRITERION VARIABLES						
.30		1.229	1.463	1.227	1.239	1.334
.60		1.275	1.535	1.258	1.294	1.546
CORRELATION* BETWEEN PREDICTOR AND CRITERION VARIABLES						
.00		1.347	2.128	1.491	1.313	1.501
.20		1.306	1.320	1.210	1.354	1.586
.40		1.102	1.049	1.026	1.132	1.249
SAMPLE SIZE**						
Small		1.344	1.692	1.394	1.430	1.640
Large		1.110	1.307	1.091	1.103	1.251

* Correlation in the matrix used to generate the population matrix analyzed in the study

** Small = 5(p+q); Large = 20 (p+q) where p is number of predictor variables, q is the number of criterion

TABLE 4a

WHERRY AND OLKIN/PRATT CORRECTION FORMULAE APPLIED TO
THE TOTAL REDUNDANCY STATISTIC

p	q	n	MEAN	POPULATION VALUE	WHERRY	OLKIN/PRATT	POPULATION-WHERRY	POPULATION-OLKIN/PRATT
9	5	70	0.184701	0.126378	0.121006	0.125397	0.005372	0.000981
9	5	70	0.231142	0.177453	0.171075	0.176349	0.006378	0.001104
9	5	70	0.443837	0.404501	0.400387	0.407954	0.004114	-0.003453
9	5	70	0.184086	0.126366	0.120343	0.124721	0.006023	0.001654
9	5	70	0.203826	0.140718	0.141625	0.146396	-0.000907	-0.005678
9	5	70	0.337680	0.291738	0.285936	0.292718	0.005802	-0.000980
9	5	70	0.194974	0.135215	0.132081	0.136679	0.003134	-0.001464
9	5	70	0.203880	0.140720	0.141683	0.146455	-0.000963	-0.005735
9	5	70	0.394360	0.358482	0.347044	0.354334	0.011438	0.004148
9	5	70	0.191851	0.135185	0.128714	0.133250	0.006471	0.001935
9	5	70	0.177144	0.117356	0.112858	0.117093	0.004498	0.000263
9	5	70	0.266410	0.216770	0.209098	0.214951	0.007672	0.001819
9	5	70	0.287683	0.185906	0.180836	0.186303	0.005070	-0.000397
9	5	70	0.329491	0.235254	0.228915	0.235071	0.006339	0.000183
9	5	70	0.563710	0.502448	0.498267	0.505953	0.004181	-0.003505
9	5	70	0.283465	0.178609	0.175985	0.181476	0.002624	-0.002767
9	5	70	0.299105	0.195848	0.193971	0.199639	0.001877	-0.003791
9	5	70	0.531641	0.488800	0.461387	0.469098	0.007413	-0.000298
9	5	70	0.286285	0.185910	0.179228	0.184670	0.006682	0.001240
9	5	70	0.289608	0.185180	0.183049	0.188551	0.002131	-0.003371
9	5	70	0.393120	0.311414	0.302088	0.309051	0.009326	0.002363
9	5	70	0.279705	0.178601	0.171661	0.176982	0.006940	0.001619
9	5	70	0.264452	0.155714	0.154120	0.159150	0.001594	-0.003436
9	5	70	0.345530	0.255294	0.247360	0.253747	0.007934	0.001547

TABLE 4b

WHERRY AND OLKIN/PRATT CORRECTION FORMULAE APPLIED TO
THE TOTAL REDUNDANCY STATISTIC

p	q	n	MEAN	POPULATION VALUE	WHERRY	OLKIN/PRATT	POPULATION-WHERRY	POPULATION-OLKIN/PRATT
9	5	280	0.141728	0.126378	0.126067	0.126935	0.000311	-0.000557
9	5	280	0.190322	0.177453	0.175547	0.176655	0.001906	0.000798
9	5	280	0.414103	0.404501	0.403412	0.405174	0.001089	-0.000673
9	5	280	0.140878	0.126366	0.125201	0.126065	0.001165	0.000301
9	5	280	0.154967	0.140718	0.139547	0.140484	0.001171	0.000234
9	5	280	0.303400	0.291738	0.290689	0.292219	0.001049	-0.000481
9	5	280	0.150665	0.135215	0.135167	0.136082	0.000048	-0.000867
9	5	280	0.154712	0.140720	0.139288	0.140223	0.001432	0.000497
9	5	280	0.370521	0.358482	0.359034	0.360727	-0.000552	-0.002245
9	5	280	0.148672	0.135185	0.133137	0.134042	0.002048	0.001143
9	5	280	0.133187	0.117356	0.117370	0.118193	-0.000014	-0.000837
9	5	280	0.229788	0.216770	0.215733	0.217010	0.001037	-0.000240
9	5	280	0.210733	0.185906	0.184424	0.185571	0.001482	0.000335
9	5	280	0.258705	0.235254	0.233996	0.235341	0.001258	-0.000087
9	5	280	0.516863	0.502448	0.500759	0.502577	0.001689	-0.000129
9	5	280	0.204755	0.178609	0.178247	0.179367	0.000362	-0.000758
9	5	280	0.219143	0.195848	0.193115	0.194299	0.002733	0.001549
9	5	280	0.484996	0.468800	0.467289	0.469643	0.000971	-0.000843
9	5	280	0.211843	0.185910	0.185571	0.186724	0.000338	-0.000814
9	5	280	0.209357	0.185180	0.183003	0.184143	0.002177	0.001037
9	5	280	0.333463	0.311414	0.311245	0.312831	0.000169	-0.001417
9	5	280	0.202607	0.178601	0.176028	0.177138	0.002573	0.001463
9	5	280	0.182422	0.155714	0.155170	0.156183	0.000544	-0.000469
9	5	280	0.278470	0.255294	0.254419	0.255837	0.000875	-0.000543